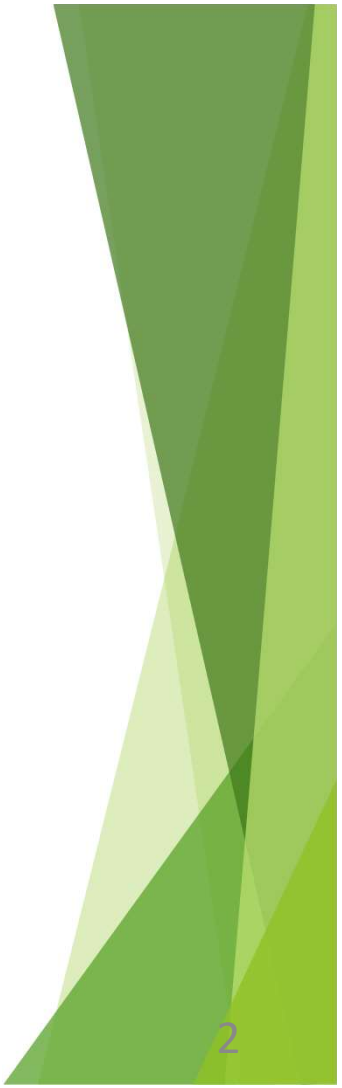
Lecture 35: Intro to Machine Learning & Ethics

COMPSCI110

Today's class

- Introduction to Machine Learning (ML)
  - What is ML?
  - Regression vs classification problems
  - Supervised vs unsupervised learning
  - A few examples
- Misuse, Ethics, biases and points of caution in ML

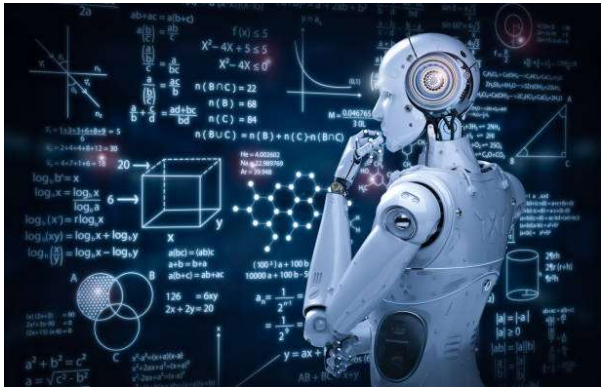


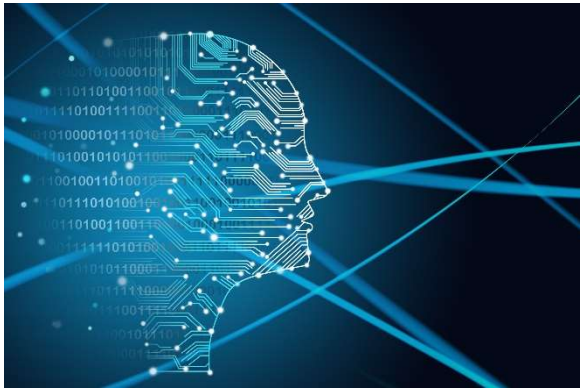
Today's class outcomes

- Introduction to Machine Learning (ML)
  - Understand the basics of how ML works, and a few challenges associated
  - Discover a few applications

- Misuse of ML and ethics considerations
  - Understand how ML can be misused
  - Know a few simple ways to be critical when using or facing it
  - Be aware of ethics considerations around ML

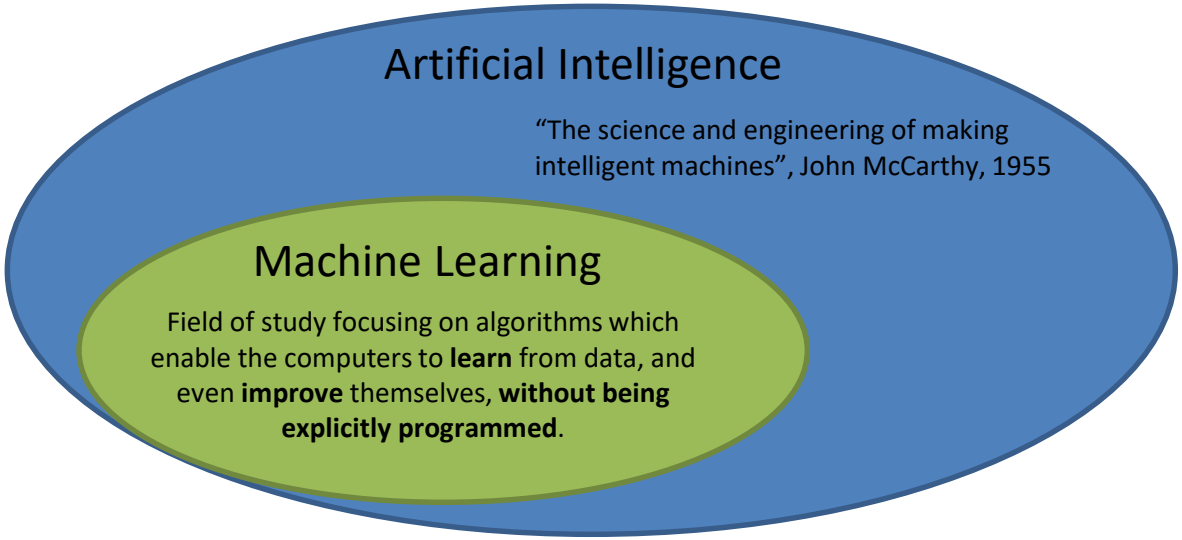
What is ML?



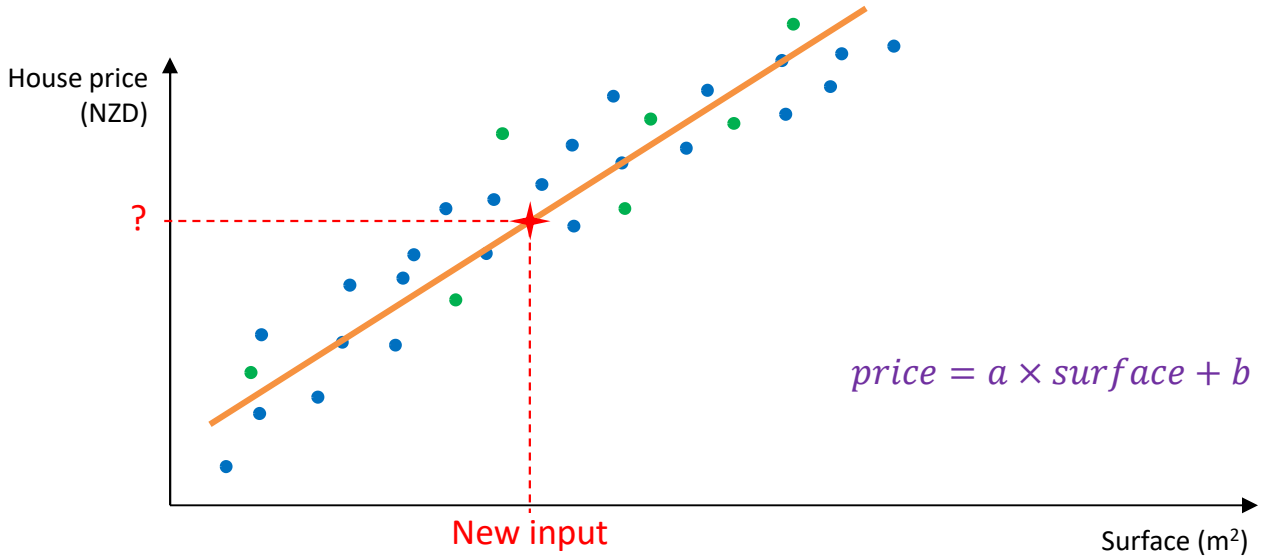


Credits : Getty Images/iStockphoto

What is ML?

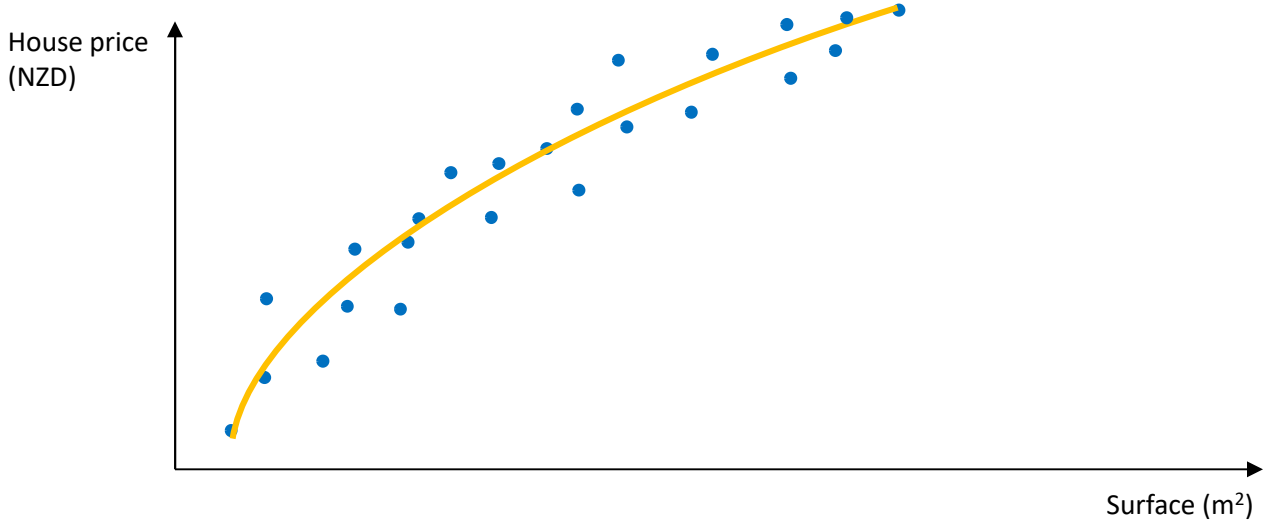

Artificial Intelligence
"The science and engineering of making intelligent machines", John McCarthy, 1955
Machine Learning
Field of study focusing on algorithms which enable the computers to **learn** from data, and even **improve** themselves, **without being explicitly programmed**.

How does ML works?



1) ML algorithms aim at uncovering *patterns* in a set of *training data* (e.g., mapping inputs and outputs).
2) A *model* is created out of the uncovered patterns.
3) The model's performance is tested with *test data*.
4) If performances are satisfactory, the model can be used to predict new inputs.

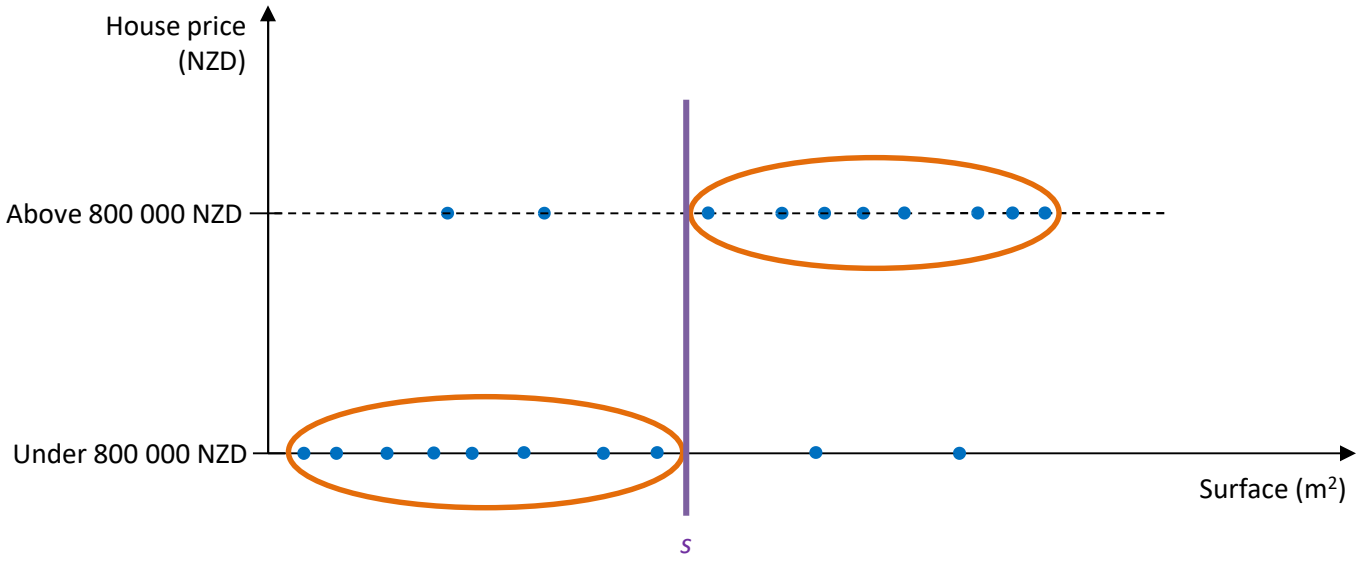
How does ML works?



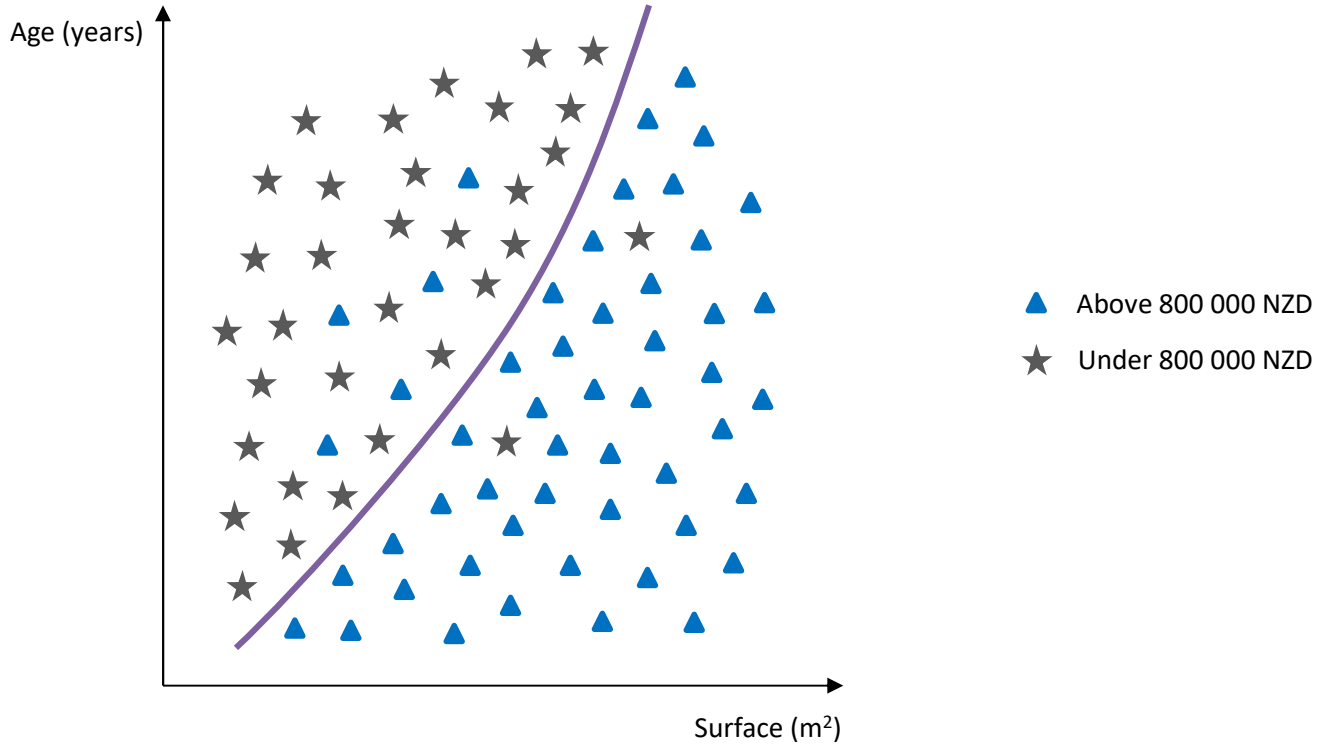
- Relationship between inputs and outputs can be linear or not.
- Several types of ML algorithms can be used to build models (e.g., Linear Regression).

Regression vs Classification

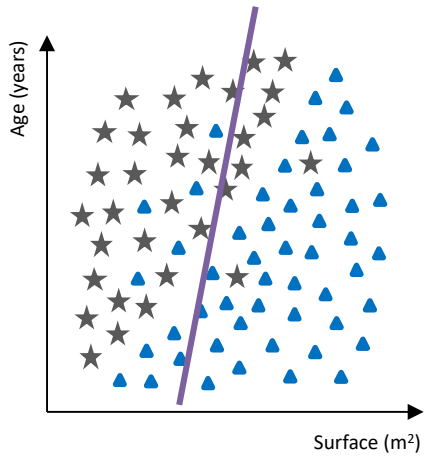
- Regression problem → predicting continuous values.
- Classification problem → predicting discrete values.



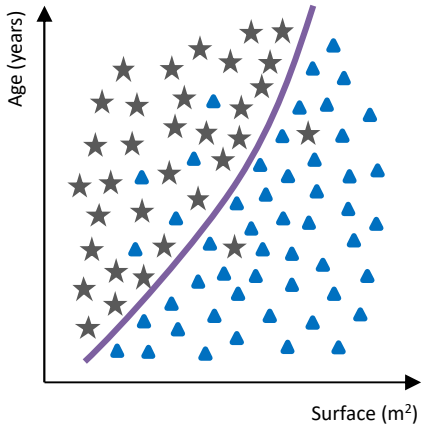
Most problems are multi-dimensional (several predictive features):



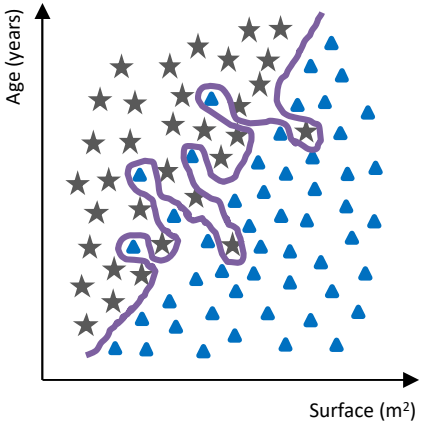
Underfitting / Overfitting



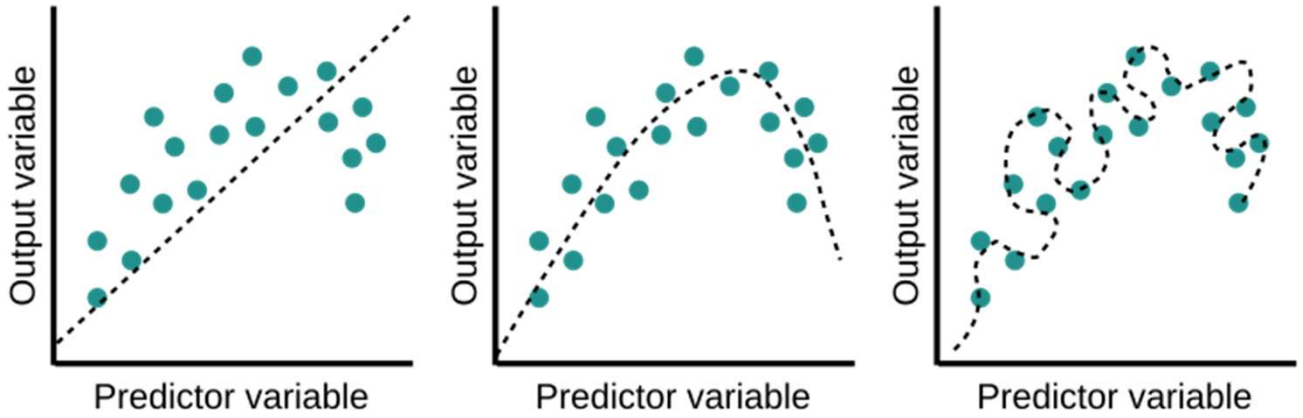
Underfitting



Good fit



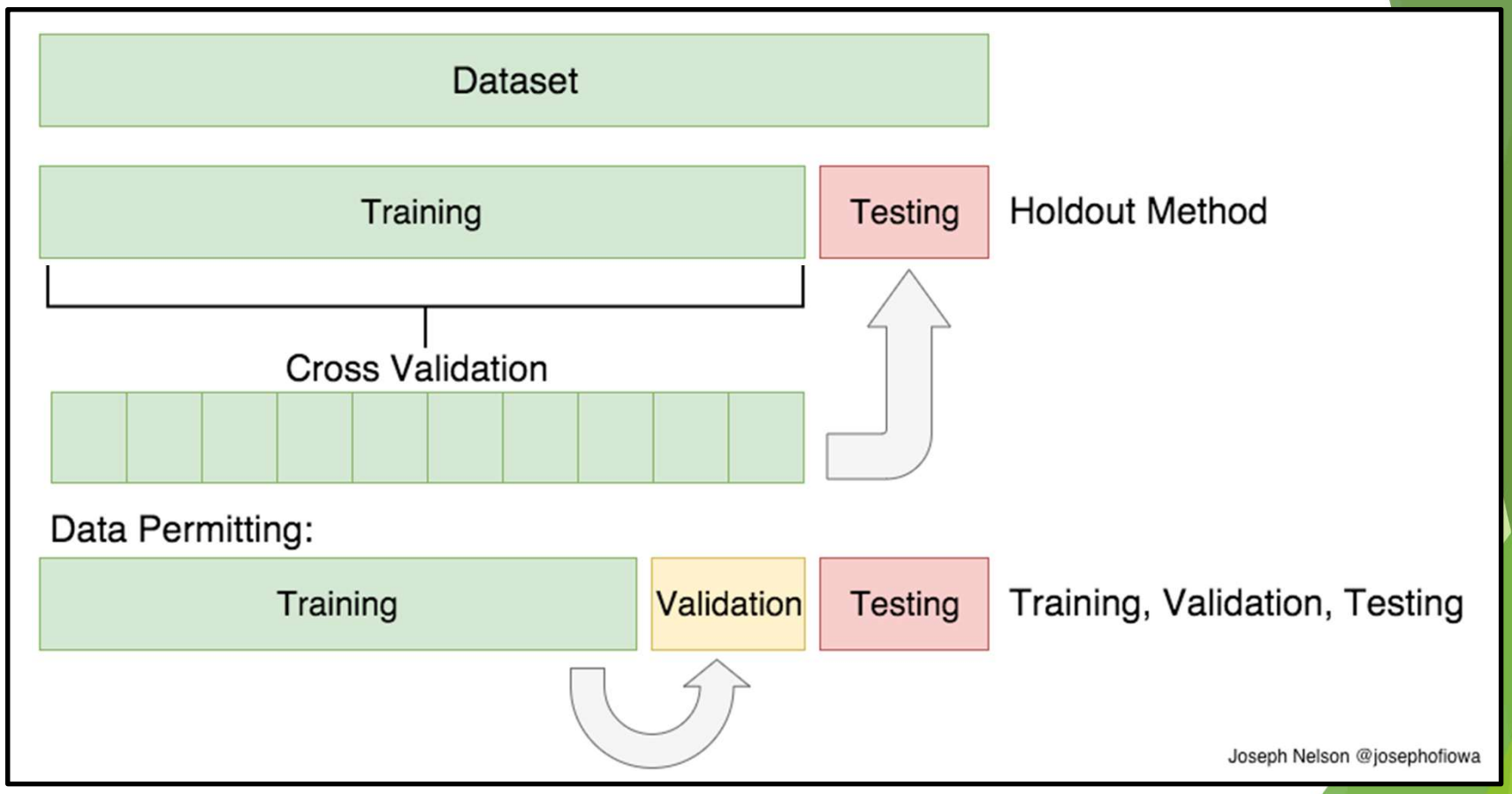
Overfitting



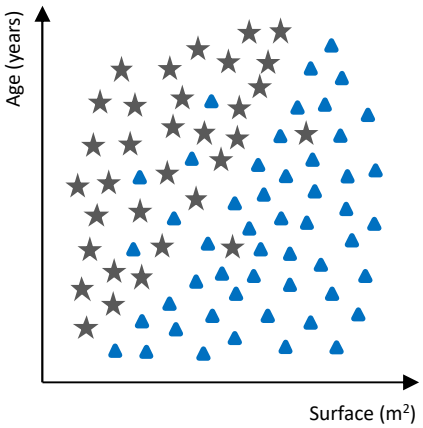
Underfitting Good fit Overfitting

- Overfitting is a modeling error that occurs when a function is too closely fit to a limited set of data points.
- Not desirable because we want model able to *generalize* (real data always contain some noise).

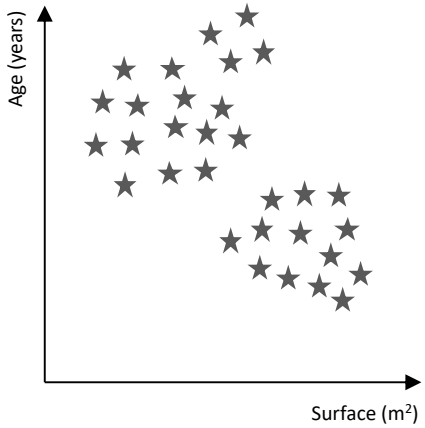
Training, Tuning and Testing Models



Supervised vs unsupervised learning

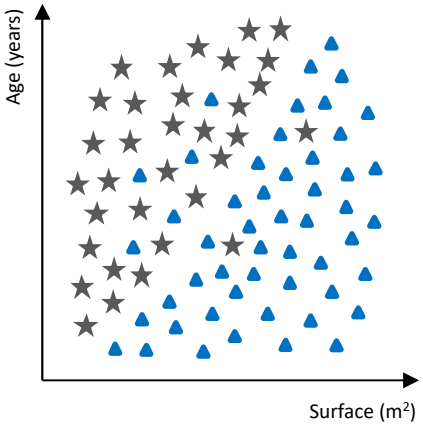


Supervised

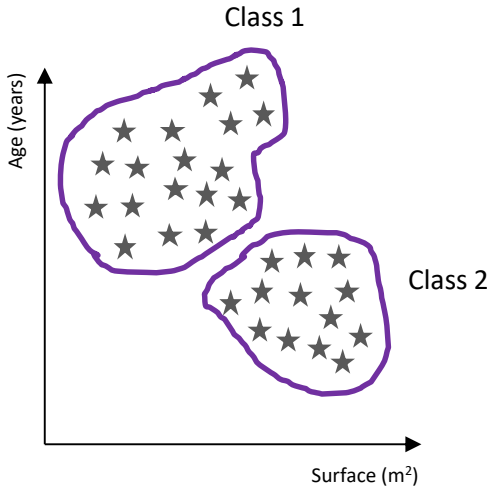


Unsupervised

Supervised vs unsupervised learning



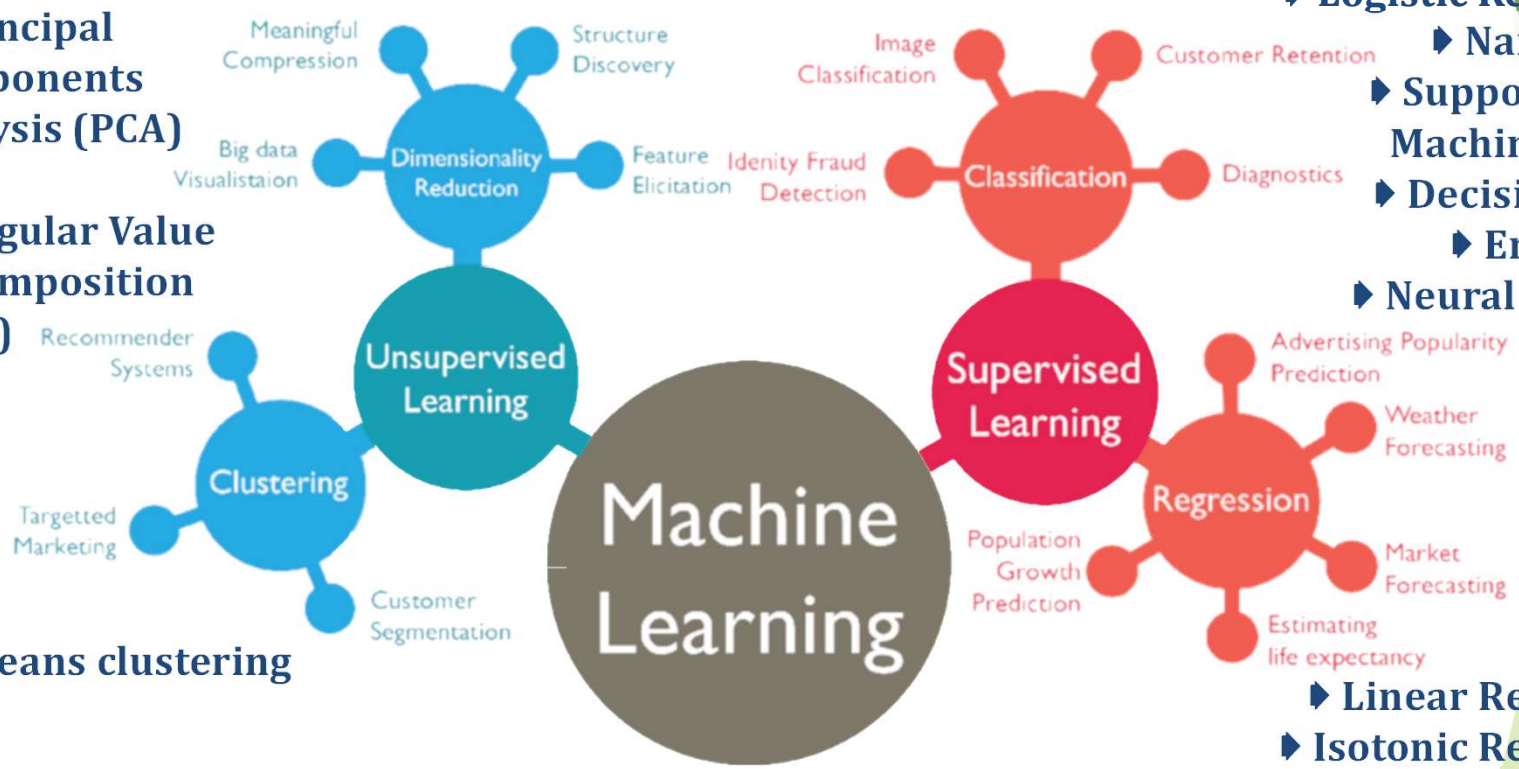
Supervised



Unsupervised

Machine Learning Taxonomy and Examples

- Principal Components Analysis (PCA)
- Singular Value Decomposition (SVD)
- k-means clustering



- Logistic Regression
- Naive Bayes
- Support Vector Machines(SVM)
- Decision Trees
- Ensembles
- Neural Network

- Linear Regression
- Isotonic Regression
- Decision Trees
- Ensembles

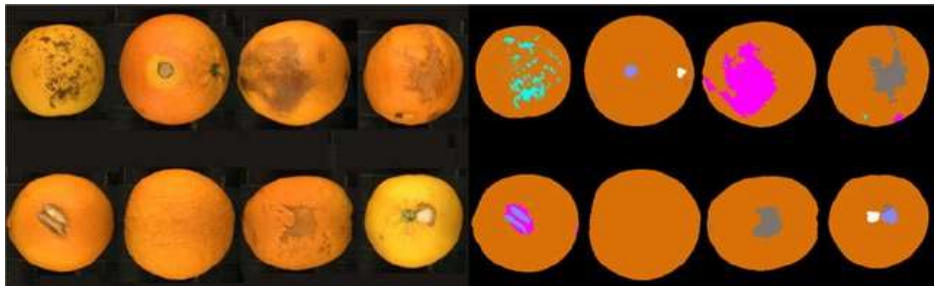
For more information, see: https://towardsdatascience.com/machine-learning-algorithms-in-laymans-terms-part-1-d0368d769a7b

A few examples

1. Industrial production: Visual quality control.
2. Medical: Detection of tumors on MRIs images.
3. Insurance: Predicting damage costs form earthquakes.
4. Environment: Predicting air quality levels.
5. Internet: Recommender systems.
6. Security: Facial recognition.

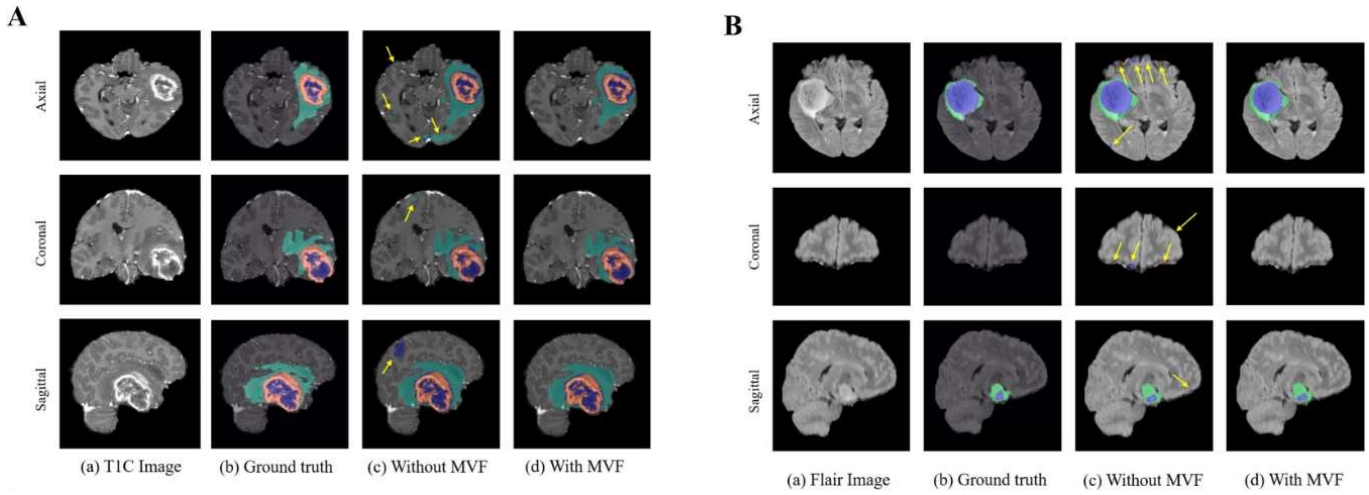
Visual Quality Control

- Machine Learning applied to computer vision.
- Example: Detecting peel defects on fruits.



Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

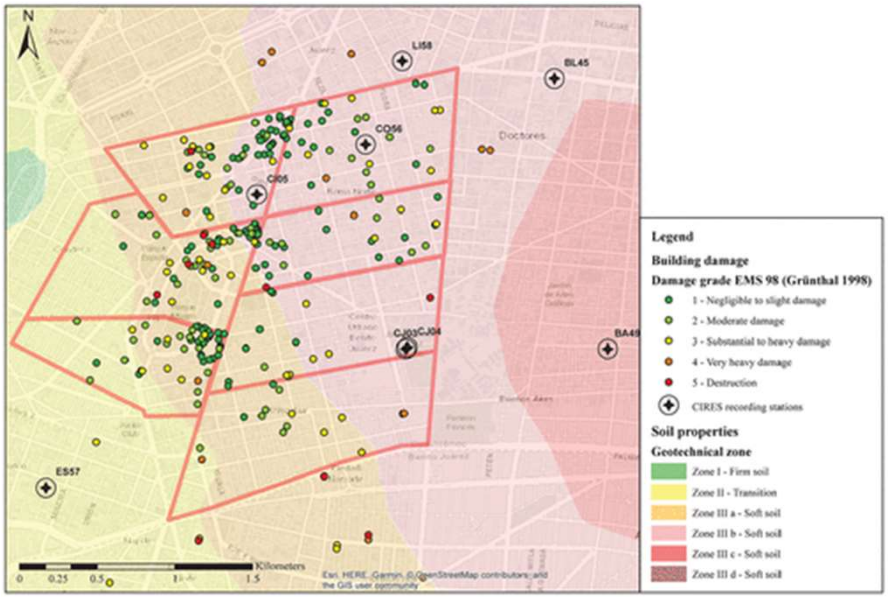
Brain tumor detection



Example Segmentation results. High-grade glioma (HGG) and (A) low-grade glioma (LGG) (B).
(a) A 2D slice of a postcontrast image,
(b) Ground truth image,
(c) Network output without multivolume fusion (MVF), and
(d) Network output with MVF.
The arrows in (c) represent false positives that are successfully eliminated after MVF (d).

Yogananda, C. G. B., Shah, B. R., Vejdani-Jahromi, M., Nalawade, S. S., Murugesan, G. K., Yu, F. F., ... & Fei, B. (2020). A Fully automated deep learning network for brain tumor segmentation. Tomography, 6(2), 186.

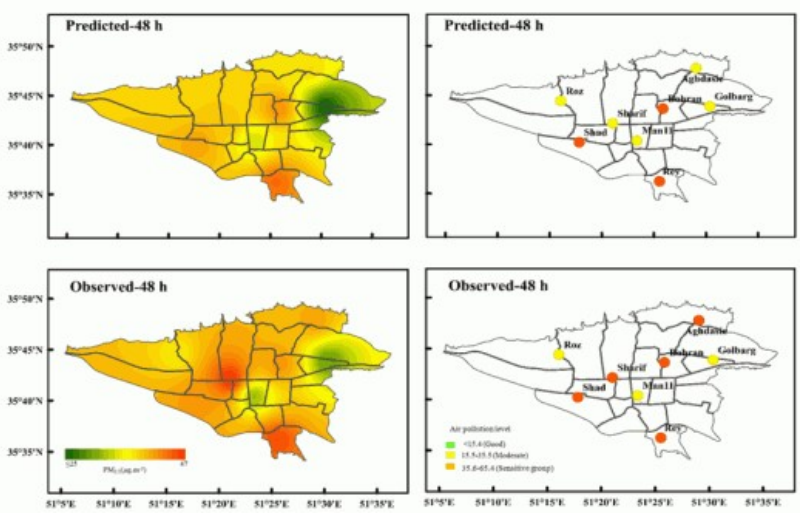
Predicting building damage



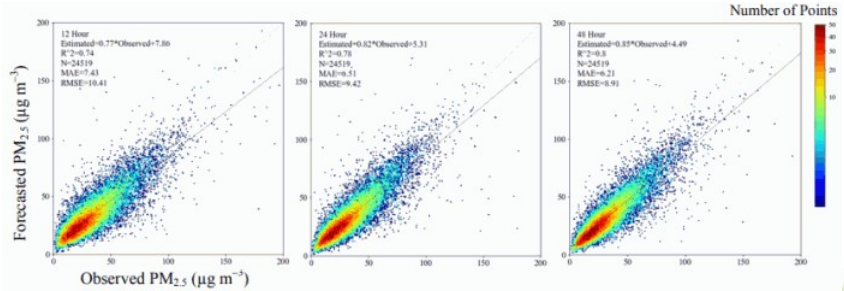
Buildings assessed in the Roma and Condesa neighborhoods, after the 2017 Puebla-Morelos, Mexico, earthquake

Roeslin, S., Ma, Q., Juárez-Garcia, H., Gómez-Bernal, A., Wicker, J., & Wotherspoon, L. (2020). A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. Earthquake Spectra, 8755293020936714.

Predicting air quality



Maps of predicted and observed fine particles concentrations PM0.5 in the region of Tehran, Iran.

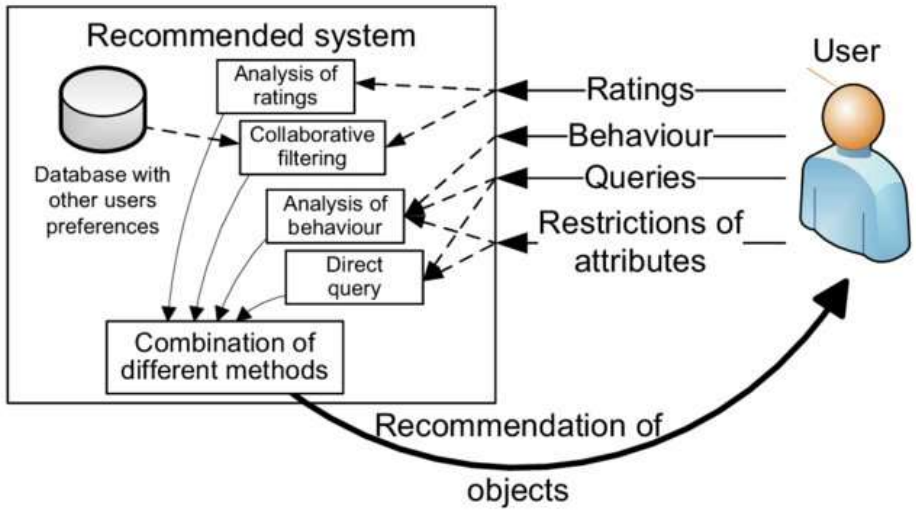


Plots of predicted vs observed values over 12, 24 and 48 hours.

Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., ... & Sachdeva, S. (2019). Evaluation of different machine learning approaches to forecasting PM2. 5 mass concentrations. Aerosol and Air Quality Research, 19(6), 1400-1410.

Recommender systems

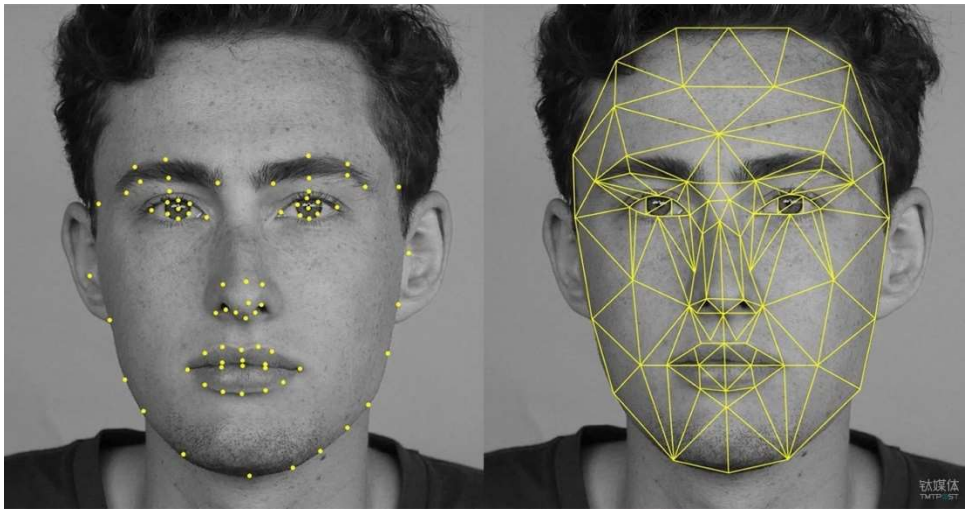
- Netflix recommends new movies/series.
- Facebook recommends new posts.
- Youtube recommends new videos.
- Etc…



Eckhardt, A. (2009). Various aspects of user preference learning and recommender systems. In DATESO (pp. 56-67).

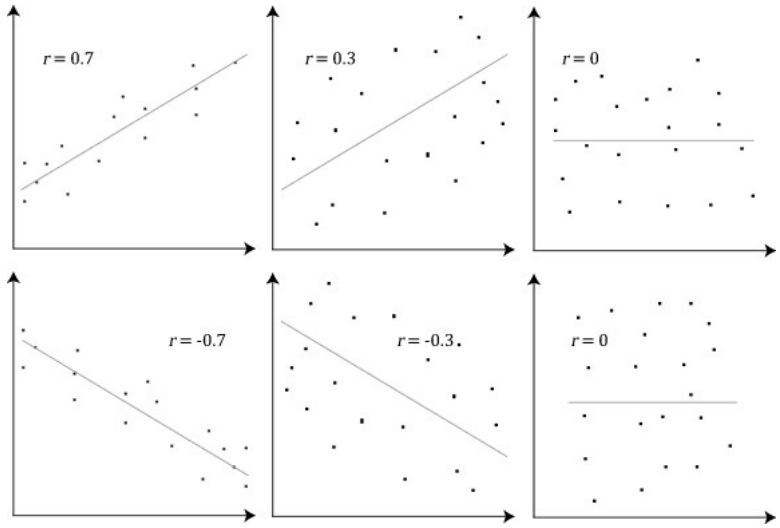
Facial recognition

- To unlock your phone/computer.
- To recognize your face before applying filters.
- To check you at a border (e.g., airport).
- To monitor people in public spaces.



Misuse of ML

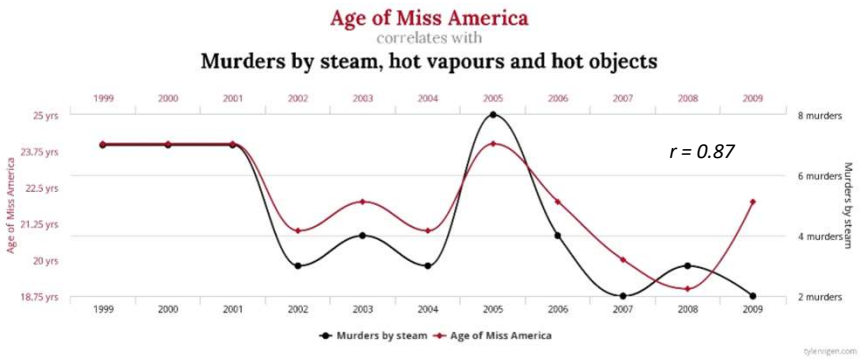
Correlation vs causality

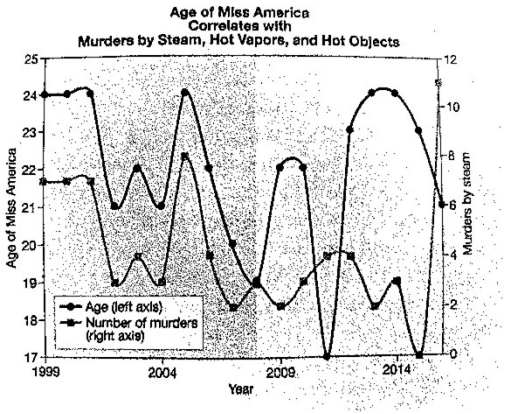


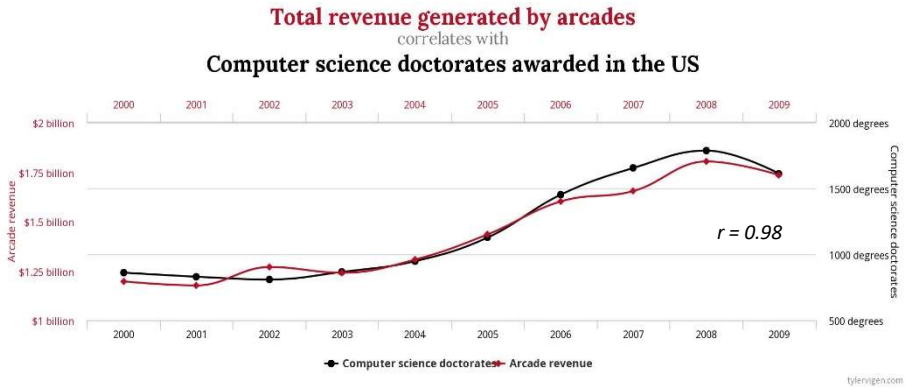
- Correlation = relation/association between the evolution of two variables.
- **Correlation does not imply causation**.
- Correlation usually refers to linear correlation, but the former holds for any type of association.
- Feels natural to infer that when two things are associated, one cause the other, but it can be misleading.

By Laerd Statistics - https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=84255641

Spurious correlations



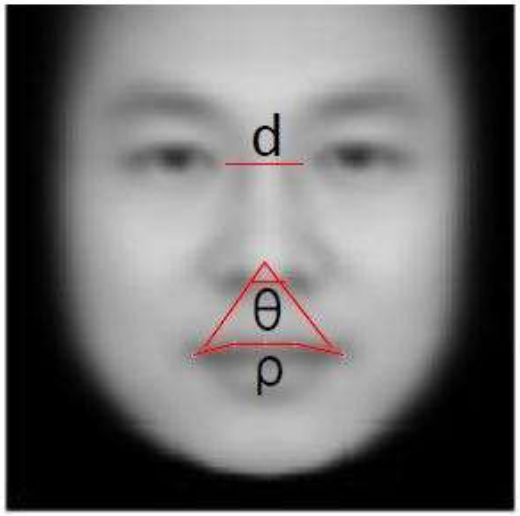




https://www.tylervigen.com/spurious-correlations

Bergstrom, C. T., & West, J. D. (2020). Calling bullshit: the art of skepticism in a data-driven world. Allen Lane.

Correlations between facial features and criminality



	Mean		Variance	
	criminal	non-criminal	criminal	non-criminal
ρ	0.5809	0.4855	0.0245	0.0187
d	0.3887	0.4118	0.0202	0.0144
θ	0.2955	0.3860	0.0185	0.0130

Table 4. The mean and variance for three normalized discriminative features ρ, d and θ.

Wu and Zhang (2016) "Automated Inference on Criminality using Face Images". https://arxiv.org/pdf/1611.04135v2.pdf

Classifier algorithm	AUC
Convolutional Neural Network (CNN)	0.9540
Support Vector Machine (SVM)	0.9303
K-Nearest Neighbour (KNN)	0.8838
Linear Regression (LR)	0.8666

Wu and Zhang (2016) "Automated Inference on Criminality using Face Images". https://arxiv.org/pdf/1611.04135v2.pdf

1) Let's take a look at the data:

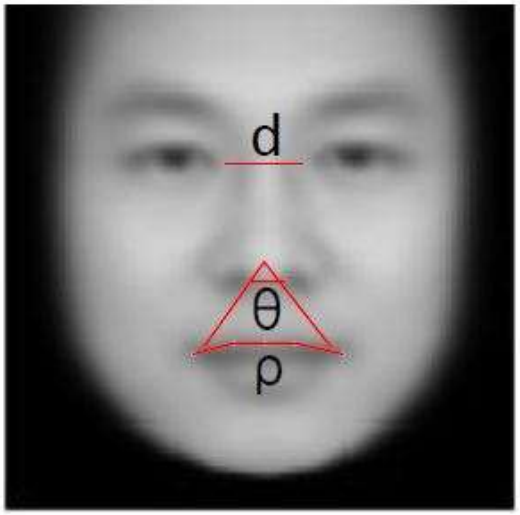
(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

Figure 1. Sample ID photos in our data set.

Wu and Zhang (2016) "Automated Inference on Criminality using Face Images". https://arxiv.org/pdf/1611.04135v2.pdf

2) Let's analyse the model output:



	Mean		Variance	
	criminal	non-criminal	criminal	non-criminal
ρ	0.5809	0.4855	0.0245	0.0187
d	0.3887	0.4118	0.0202	0.0144
θ	0.2955	0.3860	0.0185	0.0130

Table 4. The mean and variance for three normalized discriminative features ρ, d and θ.

Wu and Zhang (2016) "Automated Inference on Criminality using Face Images". https://arxiv.org/pdf/1611.04135v2.pdf

3) Flawed assumptions about “criminal type”:

i. The appearance of a person’s face is purely a function of innate properties;

ii. “Criminality” is an innate property of a group of people;

iii. Criminal judgment by a legal system reliably determines “criminality” in a way that is unaffected by facial appearance

“Machine learning does not distinguish between correlations that are causally meaningful and ones that are incidental.”

“Physiognomy’s New Clothes” Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov

"We are the first to study automated face-induced inference on criminality free of any biases of subjective judgments of human observers."

→ Bias in the data is reflected in the outcome.

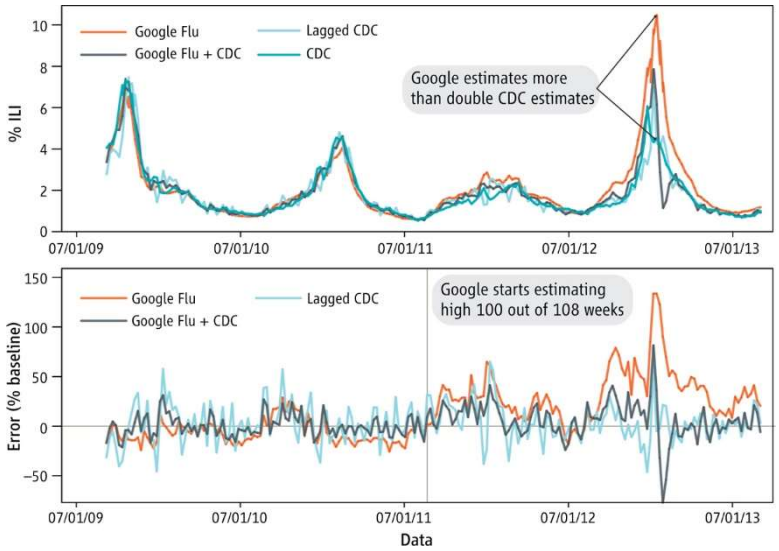
→ It is not because you use a ML algorithm that your results are objective.

Google Flu Trends

- Web service provided by Google between 2008 and 2015.
- Estimating influenza activity in real time based on Google search queries.

https://www.callingbullshit.org/

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012-1014.

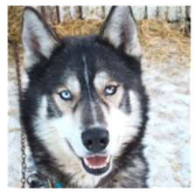


- Good real-world example of spurious correlation and overfitting.
- Some terms like “high school basketball” correlated by chance (not good for generalization).
- Search behavior changed over time (“suggest feature”).

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343(6176), 1203-1205.

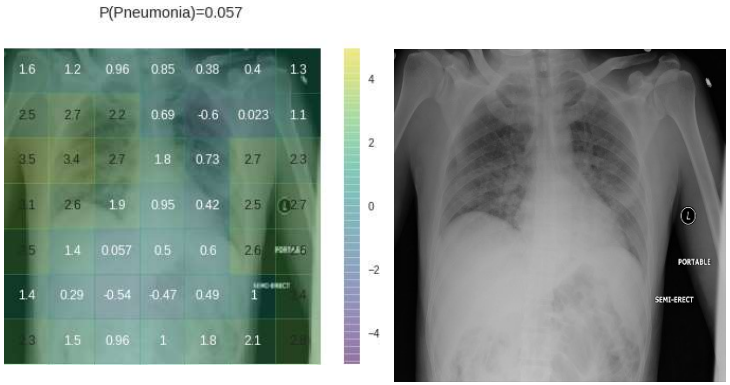
Understanding ML predictions

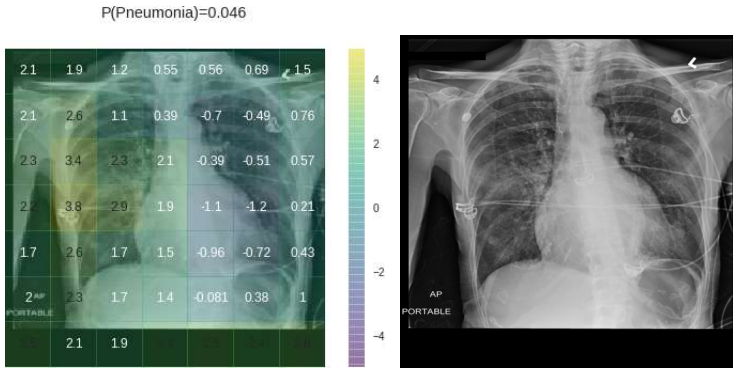
- Distinguishing wolves from huskies


(a) Husky classified as wolf


(b) Explanation

- Detecting pneumonia on radiological images





Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

What are radiological deep learning models actually learning?, John Zech, Jul.9 2018, Medium.

When using ML or reading about it, be critical of:

- The data used to train the model.
- The interpretation made from the output.
- How the model is building the output (require more knowledge about how the algorithms work)

Useful rule of the thumb:

“If it sounds too good to be true, it probably is.”

Bergstrom, C. T., & West, J. D. (2020). Calling bullshit: the art of skepticism in a data-driven world. Allen Lane.

The Truth About Algorithms | Cathy O'Neil

Free talk given by O'Neil at the RSA in London, 2017.
https://www.youtube.com/watch?v=heQzqX35c9A

O’Neil: “An algorithm is an opinion embedded in math.”

Many decisions are made in the development and implementation of a model:

- What is our definition of success? (e.g., profit vs equality promotion)
- What are acceptable proxies? (e.g., number of clicks/views/subscribing)
- Is the data appropriate? (Historical bias? Privacy issues? Inaccurate data?)

↳ If input data is biased, the output is very likely to be too (e.g., gender inequality, racism.)

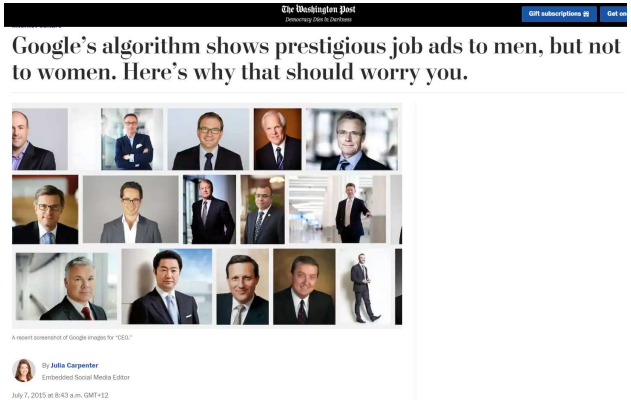
What is bias in the context of data science ethics?

- A model is considered biased in the relevant sense when one or more groups are systematically harmed by the predictions made by that model.
- More colloquial ways of saying this: “the model is unfair” to some group or “the model discriminates against…”.
- One of the most common cause of bias is that the training data is missing samples for underrepresented groups/categories (can introduce a loop reinforcement effect).

Biases in Machine Learning - Nitin Aggarwal – Towardsdatascience

The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford:
https://www.youtube.com/watch?v=fMym_BKWQzk&ab_channel=TheArtificialIntelligenceChannel

Consequences of biased ML



“Whether the data scientists behind this and other applied projects recognize it or not, their decisions about what problems to work on, what data to use, and what solutions to propose involve normative stances that affect the distribution of power, status, and rights across society. They are, in other words, engaging in political activity.”

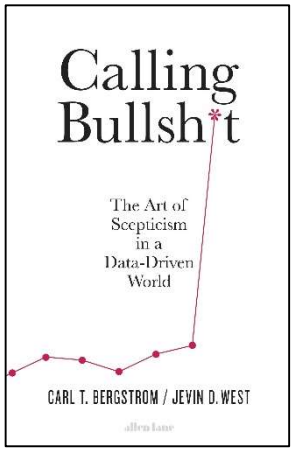
(Ben Green, “Data Science as Political Action: Grounding Data Science in a Politics of Justice” p. 22.)

Ethics and fairness tools

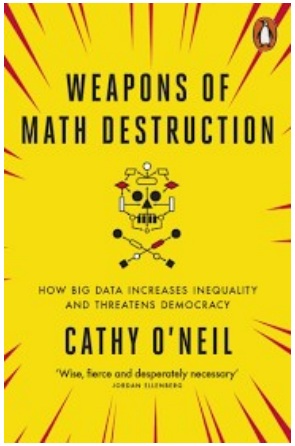
- Google's Tensorflow "fairness indicators" to detect biases in a model.
- Microsoft research FATE (fairness, accountability, transparency and ethics) group to study social implications of artificial intelligence.
- IBM open source AI fairness 360 tool to help build bias free models.
- The paper "Delayed Impact of Fair Machine Learning" by Liu et al, awarded best paper award at ICML 2018.

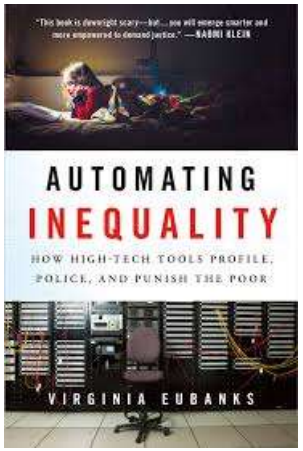
http://www.oneilrisk.com/

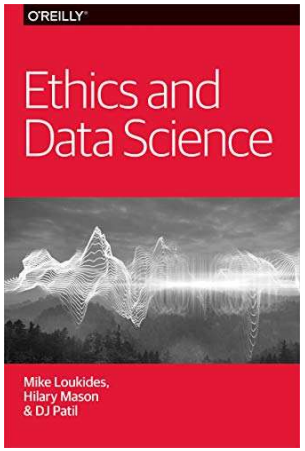
http://aequitas.dssg.io/

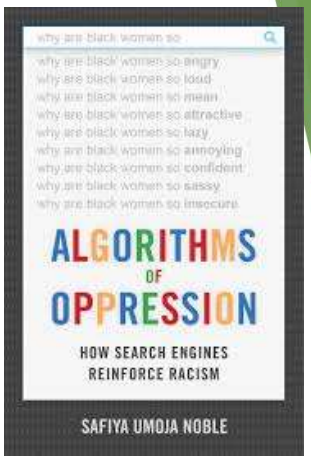


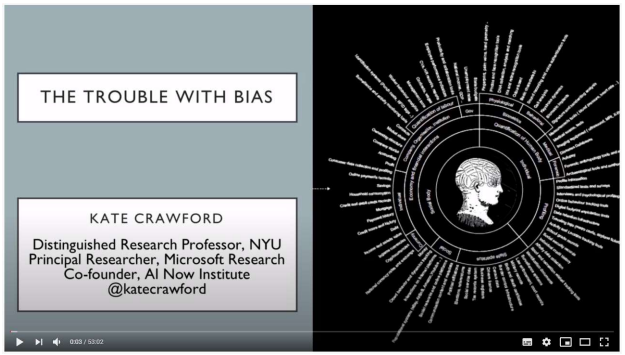
https://www.callingbullshit.org/











The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford:
https://www.youtube.com/watch?v=fMym_BKWQzk&ab_channel=TheArtificialIntelligenceChannel